DOCUMENT RESUME

ED 273 651 TM 860 500

AUTHOR Stewart, Krista J.
TITLE Development of WISC-R Administration Observational Checklist.
PUB DATE Aug 85
NOTE 12p.; Paper presented at the Annual Meeting of the American Psychological Association (Los Angeles, CA, August 23-27, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Psychology; Elementary Secondary Education; Examiners; Higher Education; *Intelligence Tests; *Interrater Reliability; *Measurement Techniques; *Observation; Participant Observation; *Rating Scales; *Testing; Testing Problems
IDENTIFIERS Wechsler Intelligence Scale for Children (Revised); *WISC R Administration Observational Checklist

ABSTRACT

The Wechsler Intelligence Scale for Children-Revised (WISC-R), one of the most commonly used tests of cognitive ability, is difficult to administer accurately. The purpose of this study was primarily to assess interrater agreement on the WISC-R Administration Observational Checklist (WAOC), a new observational instrument that can be used by an observer to evaluate all components of WISC-R administration. A secondary purpose of the study was to evaluate two WISC-R administrations of five students enrolled in a graduate course in psychoeducational assessment. Based on a total of 10 observations by two raters, Cohen's Kappa was calculated for 29 of the measures on the checklist. The values for 22 of the measures were significant (p greater than .05). The difference in mean scores for the first and second observations of the students did not quite reach statistical significance because of the small number of subjects and because of a ceiling effect for one student who scored highest on the first observation. However, after receiving feedback, students showed improvement on a number of the measures. WAOC enables the observer to pinpoint examiner errors and to give specific feedback regarding those errors. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Development of the WISC-R Administration Observational Checklist

Krista J. Stewart

Tulane University

ED273651

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. J. Stewart

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the meeting of the American Psychological Association, Los Angeles, 1985

TM 860 500

Abstract

The purpose of the present study was primarily to assess interrater agreement on the WISC-R Administration Observational Checklist (WAOC) and secondarily to evaluate two WISC-R administrations of five students enrolled in a graduate course in Psychoeducational Assessment. Based on a total of 10 observations by two raters, Cohen's κ was calculated for 29 of the measures on the checklist. The values for 22 of the measures were significant ($p > .05$). The difference in mean scores for the first and second observations of the students did not quite reach significance, though after receiving feedback, students showed improvement on a number of the measures.

Development of the WISC-R Administration Observational Checklist

Because the results of a test are only as accurate as its administration, a test must consistently be administered according to its standard directions in order to keep administration error as a source of error variance at a minimum. The Wechsler Intelligence Scale for Children-Revised (WISC-R), one of the most commonly used tests of cognitive ability, is also one of the more difficult to administer accurately. Based on observations of WISC-R administration, Fantuzzo, Sisemore, and Spradlin (1983) determined that Comprehension and Vocabulary were the most difficult Verbal subtests to administer with the major source of error being failure to accurately probe ambiguous responses. Block Design and Picture Arrangement were the most inaccurately administered Performance subtests with the major errors involving departures from standardized verbal instructions and nonstandard manipulations of test materials. Another major error was lack of adherence to standardized presentation of digits on Digit Span.

Attempts have been made within the context of assessment training to insure accuracy of Wechsler administration (Boehm, Duker, Haesloop, & White, 1974; Fantuzzo et al., 1983). Neither of these existing approaches, however, makes possible the precise pinpointing of an examiner's administration errors.

An new observational instrument, the WISC-R

Administration Observational Checklist, (WAOC, Stewart, 1984) can be used by an observer to evaluate all components of WISC-R administration, item by item. The WAOC is divided into 12 subtests each corresponding to one of the WISC-R subtests. In turn, each subtest contains either two or three sections. For each subtest the first section is used to assess an examiner's accuracy in using correct starting and stopping points, administering early test items, and following other general instructions on the test. The second and primary section for each subtest is used to evaluate the main body of the administration of the subtest. Using a "yes"/"no" format, this section is used to assess the accuracy of verbal directions, manipulation of test materials, and timing for every item administered. Appropriateness of the use of queries is also evaluated in this section. For certain subtests a third section is included to cover "special considerations" that typically occur infrequently during the course of the administration of the test (e.g., use of special prompts). To aid in the use of the checklist, exact directions from the manual are included within the context of the checklist.

The purpose of the present study was primarily to establish the interrater reliability of the WAOC. In the process of doing so, the WAOC was used to evaluate the WISC-R administration of five students enrolled in a graduate course in Psychoeducational Assessment.

Method

Subjects and Observers

The subjects in the study were five graduate students enrolled in their first of two courses in Psychoeducational Assessment. The two observers in the study were the instructor of the assessment course (also the developer of the WAOC) and the graduate teaching assistant for the course, a student already having completed the course.

Procedure

At the beginning of instruction on WISC-R administration, all students were given a copy of the WAOC and were told that their test administration would be evaluated using the checklist. The course instructor and the graduate teaching assistant observed all students on their fourth formal administration of the test (i.e., to a child in a public school setting). All observations were made simultaneously by the two observers in order to obtain interrater agreement data on the checklist. Subsequently, the instructor gave detailed feedback to each student regarding exactly where he or she had made errors and how those errors might be avoided on future administrations of the test. The students were then observed by the instructor and teaching assistant during their seventh formal administration of the test. For the second observation the students were asked to test a child of the same age as the one tested during the first observation in order to keep the

items administered as similar as possible for the two evaluations.

Results

In order to evaluate interrater agreement taking into account the proportion of agreement due to chance, Cohen's (1960) κ was calculated for each of the 29 measures used to evaluate the main body of the test administration (i.e., reading of directions, manipulations, timing, etc.). Agreement scores were based on the sum of all cases (i.e., observations on individual test items) across 10 observation sessions for each of the 29 measures.

The κ value for each of the 29 measures as well as simple proportion of agreement scores (i.e., to provide for comparison) are reported in Table 1. The κ values for 20 of the measures were significant at $p > .01$ and 2 were significant at $p > .05$. Of the 7 measures that were not significant, 5 were measures of timing. The low interrater agreement on timing was primarily due to difficulty in determining when the examiner was actually starting and stopping the watch. Agreement on the reading of digits for Digit Span was not significant because only one error was made in the 163 cases observed, and the raters were not in agreement on that case. The κ value for the reading of Vocabulary also was not significant and suggested a need for further clarification of how that subtest would be evaluated.

- - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - -

Mean scores for the first observations and the second observations based on the 29 measures were 68.4% and 83.1%, respectively. While this difference did not quite reach statistical significance, there was a trend toward significance [$t(4) = -2.34$, $p = .08$]. The fact that statistical significance was not achieved was not surprising, however, because of the small number of subjects and because of a ceiling effect for at least the one subject who scored highest on the first observation.

During the initial observations, errors were particularly common in the reading of the directions for Picture Arrangement, Block Design, and Coding; in the manipulations accompanying Picture Arrangement, Block Design, and Coding; and in the speed of reading the digits on Digit Span. In addition, the students often had difficulty with timing, apparently due to lack of practice in holding, starting, stopping, and resetting their stopwatches. Improvement was made in all of these areas by the time of the second set of observations. Errors in querying were most common on Similarities, Vocabulary, and Comprehension. While improvement was made on Similarities and Comprehension by the second observations, many errors were still occurring on Vocabulary. Errors in the cautions for Picture Completion and repetitions for Arithmetic were frequently made during

first observations but seldom occurred during the second observations.

Discussion

Interrater agreement on the various observation measures on the WAOC was generally high. In general, the data from this study support its usefulness as a training tool.

The WAOC made possible giving the students specific feedback regarding errors after the first observations. During the second observations, students showed improvement in their test administration, correcting many of the errors made during the first observations. Further research is necessary, however, to determine whether or not feedback based on the WAOC is more facilitative than less systematic, less objective feedback.

The WISC-R Administration Observational Checklist has been developed as an aid to help increase accuracy of administration of the WISC-R. This goal is based on the need to keep administration error at a minimum in order to maximize the reliability and validity of the test results. The WAOC enables the observer to pinpoint examiner errors and in turn to give specific feedback regarding those errors. Although other strategies for improving administration of the WISC-R have been attempted (e.g.,Fantuzzo, et al. 1983), the WAOC makes possible giving more precise and detailed feedback than other approaches that have been used.

References

Boehm, A. E., Duker, J., Haesloop, M. D., & White, M. A. (1974). Behavioral objectives in training for competence in the administration of individual intelligence tests. Journal of School Psychology, 12, 150-157.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Fantuzzo, J. W., Sisemore, T. A., & Spradlin, W. H. (1983). A competency-based model for teaching skills in the administration of intelligence tests. Professional Psychology, 14, 224-231.

Stewart, K. J. (1984). WISC-R administration observational checklist. Submitted for publication.

Table 1

Level of Congruence Between Two Raters for the WISC-R Administration Observational Checklist Subscales

Measure	# of Cases	Proportion of Agreement	Cohen's Kappa	Kappa Probability
Reads				
I	200	1.00	1.00	.05
S	151	.90	.77	.01
PA	61	.93	.87	.01
A	125	.94	.63	.01
BD	36	.83	.65	.01
V	225	.98	.59	.06
OA	50	.90	.72	.01
COMP	143	.96	.55	.04
CD	30	.77	.51	.01
DS	163	.99	.00	1.00
M	38	.97	.94	.01
Manipulates				
PA	61	.92	.77	.01
BD	36	.86	.70	.01
OA	50	.80	.51	.01
CD	30	.83	.66	.01
M	35	.83	.58	.01

Table 1 (cont.)

Level of Congruence Between Two Raters for the WISC-R Administration Observational Checklist Subscales

Measure	# of Cases	Proportion of Agreement	Cohen's Kappa	Kappa Probability
Starts Time				
PA	105	.83	.44	.01
A	112	.95	.70	.01
BD	80	.84	.57	.01
OA	40	.98	.79	.07
M	66	.85	.36	.11
Stops Time				
PA	105	.87	.62	.01
A	112	.92	.60	.01
BD	80	.83	.20	.34
OA	40	.80	.32	.20
M	66	.85	.52	.01
Other				
PC (time)	205	.91	.26	.17
BD (scrambles)	79	.87	.70	.01
DS (speed)	163	.74	.42	.01